CDF/MEMO/STATISTICS/PUBLIC/7117

Limit calculations in the presence of nuisance parameters. 1. Bayesian approach.

CDF Statistics Committee

Joel Heinrich, Craig Blocker, John Conway, Luc Demortier,
Louis Lyons, Giovanni Punzi, Pekka Sinervo

July 30, 2004

Abstract

We address the common problem of setting limits in the presence of uncertainties on acceptances and backgrounds. We aim to investigate several different approaches, but here describe just a Bayesian technique.

1 The problem

A very common statistical problem of relevance to Particle Physics is the extraction of an upper limit on some hypothesised process, when there are uncertainties in quantities such as the acceptance of the detector and/or the analysis procedure, the beam intensity, the estimated background, etc. These are known in statistics as nuisance parameters, or in Particle Physics as systematics. We assume that estimates of these quantities are available from subsidiary measurements.

To specify the problem in more detail, we assume that we are performing a counting experiment in which we observe n counts, and that the acceptance has been estimated as $\epsilon_0 \pm \sigma_\epsilon$ and the background as $b_0 \pm \sigma_b$. For a signal rate s, n is Poisson distributed with mean $s\epsilon + b$. Here ϵ contains factors like the intensity of the accelerator beam(s), the running time, various efficiencies, etc. It is constrained to be non-negative, but can be larger than unity.

We aim to study and compare different approaches to this problem. In general we are interested in seeing:

- Narrow intervals or low upper limits can provide tighter rejection of incorrect values, but empty or very short intervals—intervals with very small Bayesian credibility—are undesirable. Although such intervals may formally enjoy frequentist coverage, they simply lack a kind of conditional plausibility: *given* the type of measurement we are making, the resolution of the apparatus, etc., it is very unlikely that the true value of the parameter we are interested in lies in the calculated interval.

We will also look at the mean interval length, even though it is not invariant with respect to reparametrisations of the variable of interest e.g. $1/s$ or $\ln s$, or s^2, etc.[1]

- Coverage properties. This is a measure of how often the limits that we deduce would in fact include the true value of the parameter. This requires consideration of an ensemble of experiments like the one we actually performed, and hence is an essentially frequentist concept. Nevertheless, it can be applied to a Bayesian technique.

  Coverage is a property of the technique, and not of the particular limit deduced from a given measurement. It can, however, vary with the true value of the parameter, which is in general unknown in a real measurement.

  Undercoverage (i.e. the probability of containing the true value is less than the stated confidence level) is regarded by frequentists as a serious defect. Usually coverage is required for all possible values of the physical parameter.[2] In contrast, overcoverage is permissible, but the larger intervals result in less stringent tests of models that predict the value of the parameter. For measurements involving quantised data (e.g. Poisson counting), most serious methods have coverage which varies with the true value of the parameter of interest, and hence if undercoverage is to be avoided, overcoverage is inevitable.

- Bias. In the context of interval selection, this means having a larger coverage $B(s', s_{\text{true}})$ for an incorrect value of the parameter s' than for the true value s_{true}. This needs plots of coverage versus s' for different values of s_{true}. For upper limits, $B(s'_1, s_{\text{true}}) \geq B(s'_2, s_{\text{true}})$ if s'_1 is less than s'_2, so methods are necessarily biassed for low s'. Bias thus is not very interesting for upper limits. It will be discussed in later notes dealing with 2-sided intervals.

Here we investigate a Bayesian technique. We will report on and compare with other methods (e.g. Cousins–Highland, pure frequentist, profiled frequentist) in later notes.

[1]The **median** upper limit is invariant with respect to monotonic reparametrisations, but is not commonly used.

[2]The argument is that the parameter is unknown, and so we wish to have coverage, whatever its value. This ensures that, if we repeat our specific experiment many times, we should include the true value within our confidence ranges in (at least) the stated fraction of cases. This argument may, however, be over-cautious. The location of the dips in a coverage plot like that of fig. 1 occur at values which are not fixed in s, but which depend on the details of our experiment (such as the values of ϵ and b). These details vary from experiment to experiment. Thus we could achieve 'no undercoverage for the ensemble of experiments measuring the parameter s', even if the individual coverage plots did fall below the nominal coverage occasionally. Thus in some sense 'average coverage' would be sufficient (see for example reference [1]), although it is hard to quantify the exact meaning of 'average'. It should be stated that this is not the accepted position of most High Energy Physics frequentists.

2 Reminder of Bayesian approach

Before dealing with the problem of extracting and studying the limits on s as deduced from observing n events from a Poisson distribution with mean $s\epsilon + b$ in the presence of an uncertainty on ϵ, we recall the way the Bayesian approach works for the simpler problem of a counting experiment with no background and with ϵ exactly known. Then n is Poisson distributed with mean $s\epsilon$, and Bayes' Theorem[3]

$$P(B|C) = P(C|B)P(B)/P(C) \tag{1}$$

gives

$$p(s|n) = \frac{P(n|s)\pi(s)}{\int P(n|s)\pi(s)\; ds} \tag{2}$$

where $\pi(s)$ is the prior probability density for s; $p(s|n)$ is the posterior probability density function (p.d.f.) for s, given the observed n; and $P(n|s)$ is the probability of observing n, given s.

We assume a constant prior for s,[4] and that $P(n|s)$ is given by the Poisson

$$P(n|s) = e^{-s\epsilon}(s\epsilon)^n/n! \tag{3}$$

Then[5]

$$p(s|n) = \epsilon e^{-s\epsilon}(s\epsilon)^n/n! \tag{4}$$

The limit is now obtained by integrating this posterior p.d.f. for s until we achieve the required fraction β of the total integral from zero to infinity. If β is 90%, the upper limit s_{u} is given by

$$\int_0^{s_{\mathrm{u}}} p(s|n)\; ds = 0.9 \tag{5}$$

β is termed the credible or Bayesian confidence level for the limit.

For different observed n, the upper limits are shown in the last two columns of Table 1, for $b = 0$ and for $b = 3$ respectively. The Gaussian approximation for the case $b = 0$, $n = 20$, would yield $s_{\mathrm{u}} \simeq 20 + 1.28\sqrt{20} \simeq 25.7$, which is roughly comparable to the corresponding $s_{\mathrm{u}} = 27.0451$ of the Table. For $b = 0$, it coincidentally turns out that, for this particular example, the Bayesian upper limits are identical with those obtained in a frequentist calculation with the Neyman construction and a simple ordering rule (see later note on the frequentist approach to this problem). In general this is not so. Other priors sometimes used for s are $1/\sqrt{s}$ [2] or $1/s$ [3]. Having a prior peaked at smaller values of s in general results in tighter limits for a given observed n.

[3]We follow the common convention whereby lower case π's denote prior p.d.f.'s, lower case p's denote other p.d.f.'s, upper case Π's denote prior probabilities, and upper case P's denote other probabilities. Equation (1) is true for probabilities, p.d.f.'s, or mixtures depending on whether B and/or C are discrete or continuous variables.

[4]This is an assumption, not a necessity, and is in some ways unsatisfactory. (It is implausible, cannot be normalised, and creates divergences for the posterior if used with a (truncated) Gaussian prior for the acceptance ϵ.)

[5]It turns out that the sum over n of the discrete distribution (3) and the integral over s of the continuous distribution (4) are both equal to unity. This means that the probability $P(n|s)$ and the probability density $p(s|n)$ are correctly normalised.

	$\epsilon = 1 \pm 0.1$									$\epsilon = 1 \pm 0$	
n	$b = 0$	1	2	3	4	5	6	7	8	$b = 0$	3
0	2.3531	2.3531	2.3531	2.3531	2.3531	2.3531	2.3531	2.3531	2.3531	2.3026	2.3026
1	3.9868	3.3470	3.0620	2.9019	2.8000	2.7297	2.6783	2.6391	2.6083	3.8897	2.8389
2	5.4669	4.5520	3.9676	3.6026	3.3623	3.1953	3.0736	2.9816	2.9099	5.3223	3.5228
3	6.8745	5.8618	5.0463	4.4644	4.0571	3.7666	3.5534	3.3922	3.2671	6.6808	4.3624
4	8.2380	7.1964	6.2451	5.4751	4.8914	4.4569	4.1313	3.8832	3.6904	7.9936	5.3447
5	9.5714	8.5213	7.5063	6.6022	5.8579	5.2719	4.8180	4.4660	4.1904	9.2747	6.4371
6	10.8826	9.8288	8.7885	7.8047	6.9344	6.2066	5.6184	5.1499	4.7772	10.5321	7.5993
7	12.1766	11.1203	10.0703	9.0460	8.0904	7.2450	6.5289	5.9387	5.4586	11.7709	8.7958
8	13.4570	12.3984	11.3441	10.3014	9.2952	8.3635	7.5374	6.8300	6.2380	12.9947	10.0030
9	14.7261	13.6655	12.6085	11.5575	10.5247	9.5365	8.6250	7.8142	7.1136	14.2060	11.2085
10	15.9858	14.9233	13.8641	12.8090	11.7630	10.7415	9.7701	8.8758	8.0775	15.4066	12.4073
11	17.2375	16.1732	15.1121	14.0542	13.0017	11.9621	10.9525	9.9966	9.1170	16.5981	13.5983
12	18.4823	17.4163	16.3533	15.2934	14.2371	13.1881	12.1560	11.1582	10.2162	17.7816	14.7816
13	19.7210	18.6535	17.5887	16.5269	15.4682	14.4139	13.3692	12.3452	11.3588	18.9580	15.9580
14	20.9545	19.8854	18.8191	17.7554	16.6946	15.6373	14.5856	13.5459	12.5302	20.1280	17.1280
15	22.1832	21.1127	20.0448	18.9795	17.9169	16.8572	15.8014	14.7528	13.7187	21.2924	18.2924
16	23.4078	22.3359	21.2665	20.1996	19.1353	18.0737	17.0151	15.9612	14.9161	22.4516	19.4516
17	24.6286	23.5553	22.4845	21.4161	20.3502	19.2868	18.2261	17.1689	16.1172	23.6061	20.6061
18	25.8459	24.7714	23.6992	22.6294	21.5619	20.4969	19.4344	18.3747	17.3189	24.7563	21.7563
19	27.0601	25.9844	24.9109	23.8397	22.7708	21.7042	20.6400	19.5784	18.5198	25.9025	22.9025
20	28.2715	27.1946	26.1199	25.0474	23.9770	22.9090	21.8432	20.7799	19.7191	27.0451	24.0451

Table 1: Upper 90% limits for n observed events with b background and $\epsilon = 1 \pm 0.1$ ($\kappa = 100$ and $m = 99$). Also shown are limits for $b = 0$ and $b = 3$ with fixed $\epsilon = 1$.

If the whole procedure is now repeated with a background b and a flat prior, the upper limits not surprisingly decrease for increasing b at fixed n (except for the case $n = 0$ where the limits can trivially be seen to be independent of b). This is not inconsistent with the fact that the mean expected limit for a series of measurements increases with b, i.e. experiments with larger expected backgrounds have poorer sensitivity.

2.1 Coverage

Next we can investigate the frequentist coverage $C(s_{\text{true}})$[6] of this Bayesian approach. That is, we can ask what the probability is, for a given value of s_{true}, of our upper limit being larger than s_{true}, and hence being consistent with it. This is equivalent to adding up the Poisson probabilities of eqn. (3) for those values of n for which $s_{\text{u}}(n) \geq s_{\text{true}}$ i.e.

$$C(s_{\text{true}}) = \sum_{\text{relevant } n} e^{-s_{\text{true}}\epsilon} (s_{\text{true}}\epsilon)^n / n! \tag{6}$$

As s_{true} increases through any of the values of s_{u} of the last two columns of Table 1, the coverage drops sharply. The coverage is plotted in Fig. 1.

The calculation of $C(s_{\text{true}})$ can be done as follows: The identity

$$f'(x) = e^{-x} \left[\sum_{k=0}^{n-1} \frac{x^k}{k!} - \sum_{k=0}^{n} \frac{x^k}{k!} \right] = -e^{-x} \frac{x^n}{n!} \qquad \text{for} \qquad f(x) = e^{-x} \sum_{k=0}^{n} \frac{x^k}{k!}$$

[6] This is the coverage at $s = s_{\text{true}}$ when the Poisson variable is generated with $s = s_{\text{true}}$. This differs from $B(s', s_{\text{true}})$ where the coverage is checked at $s = s'$ when the generation value is s_{true}.

allows us to write (integrating $-f'(x)$)

$$\int_0^{s_{\text{true}}} p(s|n)\,ds = \int_0^{s_{\text{true}}\epsilon} e^{-x}\frac{x^n}{n!}dx = 1 - e^{-s_{\text{true}}\epsilon}\sum_{k=0}^{n}\frac{(s_{\text{true}}\epsilon)^k}{k!}$$

From this, it follows that "relevant n" is equivalent to any one of these inequalities:

$$s_{\text{u}}(n) \geq s_{\text{true}} \Rightarrow \int_0^{s_{\text{u}}} p(s|n)\,ds \geq \int_0^{s_{\text{true}}} p(s|n)\,ds \Rightarrow \beta \geq 1 - e^{-s_{\text{true}}\epsilon}\sum_{k=0}^{n}\frac{(s_{\text{true}}\epsilon)^k}{k!}$$

and our expression for the coverage becomes

$$C(s_{\text{true}}) = 1 - \sum_{n=0}^{\prime} e^{-s_{\text{true}}\epsilon}\frac{(s_{\text{true}}\epsilon)^n}{n!}$$

where $\sum'$ means "sum until the next term would cause the sum to exceed $1-\beta$". This result proves that $C(s_{\text{true}}) \geq \beta$ for all values of s_{true} in this simple example.

It is seen that the coverage starts at 100% for small s_{true}. This is because even for $n = 0$ the Bayesian upper limit will include s_{true}, and this is even more so for larger n.

Bayesian methods are supposed to achieve average coverage. This means that when the coverage is averaged over the parameter s, weighted by the prior in s, the result should agree with the nominal value β i.e.

$$\frac{\int C(s)\ \pi(s)\ ds}{\int \pi(s)\ ds} = \beta \tag{7}$$

A proof of this theorem is given in the second appendix, section 7 of this note.

For a constant prior, the region at large s tends to dominate the average, while in general we will be interested in the coverage at small s. Thus the 'average coverage' result is of academic rather than practical interest, especially for the case of a flat prior. Indeed it is possible to have a situation where the average coverage is, say, 90%, while the coverage as a function of s is always larger than or equal to 90%.

3 The actual problem

Our actual problem differs from the simple case of Section 2 in that
(a) we have a background b, assumed for the time being to be accurately known; and
(b) we have an acceptance ϵ estimated in a subsidiary experiment as $\epsilon_0 \pm \sigma_\epsilon$.

What we are going to do is to use an extended form of Bayes' Theorem to express $p(s,\epsilon|n)$ in terms of $P(n|s,\epsilon)$ and the priors for s and ϵ. The relationship is[7]

$$p(s,\epsilon|n) = \frac{P(n|s,\epsilon)\pi(s)\pi(\epsilon)}{\iint P(n|s,\epsilon)\pi(s)\pi(\epsilon)\ ds\ d\epsilon} \tag{8}$$

[7] For the case where the probabilities have a frequency ratio interpretation, this is seen from the mathematical identities

$$P(X \text{ and } Y \text{ and } Z) = \frac{N(X \text{ and } Y \text{ and } Z)}{N(Z)}\frac{N(Z)}{N_{\text{tot}}} = P(X,Y|Z)\ P(Z) \text{ and}$$

$$P(X \text{ and } Y \text{ and } Z) = \frac{N(X \text{ and } Y \text{ and } Z)}{N(X \text{ and } Y)}\frac{N(X \text{ and } Y)}{N_{\text{tot}}} = P(Z|X,Y)\ P(X,Y).$$

So with X, Y and Z identified with s, ϵ and n respectively, and with the prior for s and ϵ factorising into two separate priors for s and for ϵ, we obtain $p(s,\epsilon|n)\ P(n) = P(n|s,\epsilon)\ \pi(s)\ \pi(\epsilon)$.

To obtain the posterior p.d.f. for s, we now integrate this over ϵ:

$$p(s|n) = \int_0^\infty p(s,\epsilon|n)\ d\epsilon, \tag{9}$$

and finally we use this to set a limit on s as in eqn. (5).

The coverage for this procedure needs to be calculated as a function of s_{true} and ϵ_{true}. The average coverage theorem of the previous section must be generalized to

$$\frac{\iint C(s,\epsilon)\pi(s)\pi(\epsilon)\, ds\, d\epsilon}{\iint \pi(s)\pi(\epsilon)\, ds\, d\epsilon} = \beta$$

3.1 Priors

To implement the above procedure we need priors for s and ϵ. As in the simple example of Section 2, for simplicity we assume that the prior for s is constant. Again it is interesting to look at the way the properties of this method change as other priors for s are used.

We assume that the prior for ϵ is extracted from some subsidiary measurement $\epsilon_0 \pm \sigma_\epsilon$. We do **not** assume that this implies that our belief about ϵ_{true} is represented by a **Gaussian** distribution centred on ϵ_0, as this would give trouble with the lower end of the Gaussian extending to negative ϵ. Instead, we specify some particular form of the subsidiary experiment that provides information about ϵ, and then assume that a Bayesian analysis of this yields a posterior p.d.f. for ϵ. Slightly confusingly, this posterior from the subsidiary experiment is used as the prior for the application of Bayes' Theorem to extract the limit on s (see eqns. (8) and (9)). We aim to have this posterior/prior for ϵ having zero probability density at $\epsilon = 0$.

3.2 The subsidiary measurement

Somewhat arbitrarily, we assume that, for a true acceptance ϵ_{true}, the probability for the measured value ϵ_0 in the subsidiary experiment is given by a Poisson distribution

$$P(\epsilon_0|\epsilon_{\text{true}}) = e^{-\kappa\epsilon_{\text{true}}}\kappa^m \epsilon_{\text{true}}^m/m! \tag{10}$$

where $\epsilon_0 = (m+1)/\kappa$, $\sigma_\epsilon^2 = (m+1)/\kappa^2$ and κ is a scaling constant[8]. We interpret this as the probability for ϵ_0. This is discrete because the observable m is discrete, but the allowed values become closely spaced for large κ. For small σ_ϵ/ϵ (i.e. for large m), these probabilities approximate to a narrow Gaussian (see fig. 2).

Given our choice of probability in eqn (10), the likelihood for the parameter ϵ, given measured ϵ_0, is

$$\mathcal{L}(\epsilon|\epsilon_0) = e^{-\kappa\epsilon}\kappa^m\epsilon^m/m! \tag{11}$$

This is the same function of ϵ and ϵ_0 as eqn. (10), but now m is regarded as fixed, and ϵ is the variable. The likelihood is a continuous function of ϵ. It is compared with a Gaussian in fig 3.

[8] Here we define ϵ_0 and σ_ϵ^2 as the mean and variance of the posterior p.d.f. of eqn (12).

Finally in the Bayes approach, with the choice of a constant prior for ϵ, the posterior probability density for ϵ after our subsidiary measurement is

$$\frac{dp}{d\epsilon} \propto \; e^{-\kappa\epsilon}\kappa^m\epsilon^m/m! \tag{12}$$

which is obtained by multiplying the right-hand side of eqn (11) by unity. This **posterior** probability density for ϵ will be used as our **prior** for ϵ in the next step of deducing the limit for s.

4 Results

The details of the necessary analytical calculations[9] are presented in the Appendix of this note. In this section we investigate the behavior of the Bayesian limits in this example, especially the shape of the frequentist coverage probability as a function of s_{true}.

4.1 Shape of the posterior

The posterior p.d.f. for s has the form

$$p(s|b,n)ds \propto \left[\int_0^\infty \frac{e^{-(\epsilon s+b)}(\epsilon s+b)^n}{n!}\frac{\kappa(\kappa\epsilon)^m e^{-\kappa\epsilon}d\epsilon}{\Gamma(m+1)}\right] ds$$

where the likelihood, the prior for ϵ, the (constant) prior for s, and the marginalization integral over ϵ are all prominently displayed.

The posterior probability density for s gives the complete summary of the outcome of the measurement in the Bayesian approach. It is therefore important to understand its shape before proceeding to use it to compute a limit (or extract a central value and error-bars).

Figure 4 illustrates the shape of the posterior for s (i.e. marginalized over ϵ) in the case of a nominal 10% uncertainty on ϵ, and an expectation of 3 background events. Plots are shown for 1, 3, 5, and 10 observed events. The posterior evolves gracefully from being strongly peaked at $s = 0$ to a roughly Gaussian shape that excludes the neighborhood near $s = 0$ with high probability. Technically, the posterior would be described as a mixture of $n+1$ Beta distributions of the 2nd kind[10], giving it a tail at high s which is more extensive than that of a Gaussian.

4.2 Upper limits

In this note, our main goal is to obtain a Bayesian upper limit s_{u} from our observation of n events. It is by integrating the posterior p.d.f. out to $s = s_{\mathrm{u}}$ that an upper limit is

[9] This example can be handled analytically. More complicated cases might require numerical integration, which can be done via numerical quadrature or Monte Carlo methods.

[10] The 2nd Beta distribution is also known as "Beta′" (i.e. "Beta prime"), "inverted Beta", "Pearson Type VI", "Variance-Ratio", "Gamma-Gamma", "F", "Snedecor", "Fisher-Snedecor"....

calculated: a $\beta = 90\%$ upper limit is defined so that the integral of the posterior from $s = 0$ to $s = s_{\mathrm{u}}$ is 0.9. The probability (in the Bayesian sense) of $s_{\mathrm{true}} < s_{\mathrm{u}}$ is then exactly β.

Table 1 shows the upper limits ($\beta = 0.9$) for $n = 0$–20 observed events with $b = 0$–8 and $\epsilon = 1 \pm 0.1$. (Integer values of b are chosen for illustration purposes only; b can, of course, take any real value ≥ 0.)

One notices that when $n = 0$, the limit is independent of the expected background b. This is required in the Bayesian approach: we know that exactly zero background events were produced (when no events at all were produced), and this knowledge of what *did* happen makes what might have happened superfluous. An interesting corollary is, in the case of no events observed, uncertainties in estimating the background rate are of no consequence in the Bayesian approach, and must not contribute any systematic uncertainty to the limit. This reasoning does not hold in the frequentist framework, where what might have happened definitely does influence the limit.

For comparison, limits for fixed $\epsilon = 1$ with $b = 0$ or $b = 3$ are also shown in Table 1. It is interesting that these two columns start out equal at $n = 0$ and differ by almost exactly 3 for $n > 11$. In contrast, the difference between the $b = 0$ and $b = 3$ columns for $\epsilon = 1 \pm 0.1$ is already greater than 3 at $n = 6$, and continues to grow as n increases; it is not clear whether the difference approaches a finite value as $n \to \infty$. In any case, the limits for $\epsilon = 1$ exactly are all smaller than the corresponding limits for $\epsilon = 1 \pm 0.1$, as expected.

4.3 Coverage

The main quantity of interest in this subsection is the frequentist coverage probability C as a function of s_{true} (for fixed ϵ_{true} and b). Because both the main and the subsidiary measurements involve observing a discrete number of events, the function $C(s_{\mathrm{true}})$ will have many discontinuities. On the other hand, $C(\epsilon_{\mathrm{true}})$ will be continuous (for fixed s_{true}). The explanation of this effect is as follows:

> The measured data are n events in the main measurement and m events in the subsidiary measurement. For each observed outcome (n, m) there is a limit $s_{\mathrm{u}}(n, m)$. This limit includes the effect of marginalization over ϵ.
>
> All (n, m) with $n \geq 0$ and $m \geq 0$ are possible, and the probability P of observing (n, m) can be calculated as the product of two Poissons. (It will depend on s_{true}, ϵ_{true}, ...) If we look at all the possible limits we can obtain,
>
> $$\{s_{\mathrm{u}}(n, m) | n \geq 0 \text{ and } m \geq 0\}$$
>
> and sort them in increasing s_{u}, the s_{u} are countably infinite in number and dense in the same way that rational numbers are dense in the reals.
>
> To compute the coverage as a function of s_{true}, we simply add up all the probabilities of obtaining (n, m) with $s_{\mathrm{u}}(n, m) \geq s_{\mathrm{true}}$:
>
> $$C = \sum_{(n,m)\in\mathcal{A}} P(n, m) \qquad \mathcal{A} = \{(n, m) | s_{\mathrm{true}} \leq s_{\mathrm{u}}(n, m) \text{ and } n \geq 0 \text{ and } m \geq 0\}$$
>
> This sum is over a countably infinite number of terms. If we increase s_{true} slightly to $s_{\mathrm{true}} + ds$ and recalculate the coverage, we have to drop all the terms
>
> $$\{(n, m) | s_{\mathrm{true}} \leq s_{\mathrm{u}}(n, m) \leq s_{\mathrm{true}} + ds\}$$

from the previous sum (the $P(n,m)$ for each term also changes continuously with $s_{\rm true}$, but this is no problem). If there are $M > 0$ such terms, there are M discontinuities in the coverage in the interval $[s_{\rm true}, s_{\rm true} + ds]$, since $P(n,m)$ for each of these is finite, and we lose them one by one as we sweep across the interval $[s_{\rm true}, s_{\rm true} + ds]$.

But it seems that, in general, we can always find a solution to $s_{\rm true} \le s_{\rm u}(n,m) \le s_{\rm true} + ds$ for finite ds by going out to larger and larger n and m. So, although the discontinuity may be tiny, we can always find a finite discontinuity in any finite interval of $s_{\rm true}$.

On the other hand, if we keep $s_{\rm true}$ fixed and vary $\epsilon_{\rm true}$, we always sum over the same set of (n,m), since the definition of $\mathcal{A}$ does not involve $\epsilon_{\rm true}$, and $P(n,m)$ is continuous in $\epsilon_{\rm true}$. So the coverage is continuous as a function of $\epsilon_{\rm true}$ for $s_{\rm true}$ fixed.

Plotting a curve that is discontinuous at every point is somewhat problematical. The solution adopted here is to plot the coverage as straight line segments between the discontinuities, ignoring any discontinuities with $|\Delta C| < 10^{-4}$. Figure 5 shows $C(s_{\rm true})$ for the case $\beta = 90\%$, $\epsilon_{\rm true} = 1$, nominal 10% uncertainty of the subsidiary measurement of ϵ, and $b = 3$. We observe that $C(s_{\rm true}) > \beta$ in this range, and it is not clear numerically whether $C(s_{\rm true}) \to \beta$ as $s_{\rm true} \to \infty$. The same conclusions hold for Fig. 6, which illustrates the same situation with a 20% nominal uncertainty for the ϵ-measurement.

Figure 7 shows $C(\epsilon_{\rm true})$ for $\beta = 90\%$, $s_{\rm true} = 10$, $\kappa = 100$, and $b = 3$—continuous as advertised. The shape of the curve is quite similar to that of Figs. 5 and 6, so it seems that the coverage probability (with b fixed) is approximately a function of just the product of $\epsilon_{\rm true}$ and $s_{\rm true}$. This approximate rule is likely to fail in the limit as $\epsilon_{\rm true} \to 0$ and $s_{\rm true} \to \infty$, for example, but it seems to hold when $\epsilon_{\rm true}$ and $s_{\rm true}$ are at least of the same order of magnitude.

When $\epsilon_{\rm true} s_{\rm true}$ is small, of order 1 or less, the coverage is ~100%, as in the simple case of Fig. 1. Otherwise, the behavior of coverage in Figs. 5–7 is superior to that of Fig. 1, which has a much larger amplitude of oscillation.

Another frequentist quantity that characterizes the performance of a limit scheme is the sensitivity, defined as the mean of $s_{\rm u}$. Figure 8 shows the sensitivity as a function of $s_{\rm true}$ for the case of Fig. 5; $\langle s_{\rm u} \rangle$ is observed to be nearly linearly dependent on $s_{\rm true}$. There is one complication here: when the subsidiary measurement observes $m = 0$ events, and the prior for s is flat, $s_{\rm u} = \infty$. Since the Poisson probability of obtaining $m = 0$ is always finite, $\langle s_{\rm u} \rangle$ is consequently infinite. So we must exclude the $m = 0$ case from the definition of $\langle s_{\rm u} \rangle$. (In Fig. 8 the probability of obtaining $m = 0$ is $e^{-100} \simeq 4 \times 10^{-44}$.)

4.4 Other priors for s

A weakness of the Bayesian approach is that there is no universally accepted method to obtain a unique "non-informative" or "objective" prior p.d.f. Reference [2], for example, states:

> Put bluntly: data cannot ever speak entirely for themselves; every prior specification has *some* informative posterior or predictive implications; and "vague" is itself much too vague an idea to be useful. There is no "objective" prior that represents ignorance.

Nevertheless, Ref. [2] does derive a $1/\sqrt{s}$ "reference prior" for the simple Poisson case, which is claimed to have "a minimal effect, relative to the data, on the final inference". This is to be considered a "*default* option when there are insufficient resources for detailed elicitation of actual prior knowledge".

Reference [3] attempts to discover the optimal form for prior ignorance by considering the behavior of the prior under reparameterizations. For the case in question, the form $1/s$ clearly has the best properties in this respect.

We are using an s^0 prior for this study, which seems to be the most popular choice, but the Appendix works out the form of the posterior using an $s^{\alpha-1}$ prior, so we can briefly here summarize the results for the $1/s$ and $1/\sqrt{s}$ cases:

The $1/s$ prior leads to an unnormalizable posterior for all observed n when $b > 0$. The posterior becomes a δ-function at $s = 0$, $s_{\mathrm{u}} = 0$ for any β, and the coverage is consequently zero for all $s_{\mathrm{true}} > 0$. This clearly is a disaster.

The $1/\sqrt{s}$ prior results in a posterior p.d.f. qualitatively similar in shape to those of Fig 4, except that the p.d.f. is always infinite at $s = 0$. For $n \gg b$, this produces an extremely thin "spike" at $s = 0$, which has a negligible contribution to the integral of the posterior p.d.f. A more significant difference (for frequentists) between the $1/\sqrt{s}$ and the s^0 case is that the coverage probability is significantly reduced: for the case of Fig. 5 the $1/\sqrt{s}$ prior pushes the minimum coverage down to $\sim$0.87. So the $1/\sqrt{s}$ prior leads to violation of the frequentist coverage requirement; it undercovers for some values of s_{true}.

From the practical point of view—trying to upset as few people as possible—the s^0 prior for the case considered in this note seems more universally acceptable than the $1/\sqrt{s}$ prior, which is objectionable from a frequentist point of view. One might also seek to further improve the coverage properties by adopting an intermediate prior. For example, an $s^{-0.25}$ prior would reduce the level of overcoverage obtained with the s^0 prior. How acceptable this approach would be within the Bayesian Statistical community is an interesting question.

It should be noted that all the prior p.d.f.'s considered in this note are "improper priors"—they cannot be correctly normalized: In the case of the s^0 and $1/\sqrt{s}$ priors, the integral from 0 to any value s_0 is finite, while the integral from s_0 to infinity is infinite. The corresponding integrals of the $1/s$ prior are infinite on both sides for all $s_0 > 0$. Improper priors are dangerous but often useful; "improper posteriors" are generally pathological. Extra care must be taken when employing improper priors to verify the normalizability of the resulting posterior—when using a numerical method to obtain the posterior, it is very easy to miss the fact that its integral is infinite.

4.5 Restrictions

We summarize here the restrictions forced on the priors for s and ϵ—see the Appendix for the analytical causes. The discussion below assumes $b > 0$. The prior for s being of the form $s^{\alpha-1}$, we must require $\alpha > 0$, as discussed above.

As specified in this note, the prior for ϵ, being taken from the posterior from the subsidiary measurement with a flat prior, has been given no freedom. Should the subsidiary measurement observe $m = 0$ events, the posterior for s is not normalizable when $\alpha \geq 1$: $s_u = \infty$ when $m = 0$ and $\alpha \geq 1$.

This behavior is due to a well known effect: the ϵ prior becomes $\kappa e^{-\kappa\epsilon}$ when $m = 0$, which remains finite as $\epsilon \to 0$. All such cases[11] yield $s_u = \infty$ when $\alpha \geq 1$; any positive $\alpha < 1$ cuts off the posterior at large s sufficiently rapidly to render it normalizable. From this point of view, a $1/\sqrt{s}$ prior may seem preferable, but on the other hand, having $s_u = \infty$ when $m = 0$ seems intuitively reasonable. (In general, we have $s_u = \infty$ for $m \leq \alpha - 1$, but $\alpha \geq 2$ are not popular choices.)

There is another approach possible to the gamma prior for ϵ: we may simply specify by fiat the form of the prior as

$$p(\epsilon|\mu)d\epsilon = \frac{\kappa(\kappa\epsilon)^{\mu-1}e^{-\kappa\epsilon}}{\Gamma(\mu)}d\epsilon$$

where μ is no longer required to be an integer. In practice, one then might obtain μ and κ from a subsidiary measurement whose result is approximated by the gamma distribution. In such cases, one must require $\mu > \alpha$ to keep the posterior normalizable. Note that in this form, μ/κ is the mean of the ϵ prior, $(\mu - 1)/\kappa$ is the mode, and μ/κ^2 is the variance. It is then important to obtain μ and κ in a consistent way from the information supplied by the subsidiary measurement. If ϵ, for example, were estimated by a maximum likelihood method, one would identify the estimate with $(\mu - 1)/\kappa$ rather than μ/κ.

5 Conclusions

Results have been presented on the performance of a purely Bayesian approach to the issue of setting upper limits on the rate of a process, when n events have been observed in a situation where the expected background is b and where the efficiency/acceptance factor $\epsilon \pm \sigma_\epsilon$ has been determined in a subsidiary experiment. Plots of the expected sensitivity of such a measurement and of the coverage of the upper limits are given. It will be interesting to compare these with the corresponding plots for other methods of extracting upper limits, to be given in future notes.

6 Appendix A—Analytical Details

Here we present the details of the analytical calculation of the posterior p.d.f. for s. For generality, we work through the calculation with a $s^{\alpha-1}$ prior; a flat prior is then the special case $\alpha = 1$.

6.1 Posterior for s with ϵ and b fixed

We measure n events from a process with Poisson rate $\epsilon s + b$, and we want the Bayesian posterior for s, given improper prior $s^{\alpha-1}$. We compute the posterior for fixed ϵ and b

[11] A Gaussian truncated at $\epsilon = 0$ is the standard example.

in this subsection; the calculation with our prior for ϵ follows in the next subsection. We have

$$\text{posterior:}\quad p(s|\epsilon,b,n)ds = \frac{1}{\mathcal{N}_s} e^{-\epsilon s}(\epsilon s + b)^n s^{\alpha-1} ds$$

where all factors not depending on s have already been absorbed into the normalization constant $\mathcal{N}_s$, which is defined by

$$\mathcal{N}_s = \int_0^\infty e^{-\epsilon s}(\epsilon s + b)^n s^{\alpha-1} ds = \frac{b^{n+\alpha}}{\epsilon^\alpha} \int_0^\infty e^{-bu} u^{\alpha-1}(1+u)^n du \qquad (u = s\epsilon/b)$$

where we have performed the indicated change of variable.

Expanding $(1+u)^n$ in powers of u using the binomial theorem, we get

$$(1+u)^n = n! \sum_{k=0}^n \frac{u^{n-k}}{(n-k)!k!} \qquad \Rightarrow \qquad \mathcal{N}_s = n!\epsilon^{-\alpha} \sum_{k=0}^n \frac{\Gamma(\alpha+n-k)b^k}{(n-k)!k!}$$

Recognizing this as of the general hypergeometric form, we write it as

$$\mathcal{N}_s = \epsilon^{-\alpha}\Gamma(\alpha+n)\left[1 + \frac{n}{\alpha+n-1}\frac{b}{1!} + \frac{n(n-1)}{(\alpha+n-1)(\alpha+n-2)}\frac{b^2}{2!} + \cdots\right]$$

to make the hypergeometric nature more explicit. Using the modern notation[4]

$$\text{falling factorial:} \qquad z^{\underline{k}} \equiv \frac{\Gamma(z+1)}{\Gamma(z-k+1)} = z(z-1)(z-2)\cdots(z-k+1)$$

this is expressed as

$$\mathcal{N}_s = \epsilon^{-\alpha}\Gamma(\alpha+n) \sum_{k=0}^n \frac{n^{\underline{k}}}{(\alpha+n-1)^{\underline{k}}} \frac{b^k}{k!} = \epsilon^{-\alpha}\Gamma(\alpha+n) M(-n, 1-n-\alpha, b)$$

where M is the notation of [5] (M, a confluent hypergeometric function, is often written ${}_1F_1$, and the relation given here is only valid for integer $n \geq 0$.) Note that $M(-n, 1-n-\alpha, b)$ is a polynomial of order n in b (for n a non-negative integer), and is related to the Laguerre polynomials. When $\alpha = 1$, we get $M(-n,-n,b)$, which is related to the Incomplete Gamma Function. When $\alpha = 0$, we get $M(-n, 1-n, b)$, which is infinite, so we require that $\alpha > 0$.

Our posterior probability density for fixed ϵ is then given by

$$p(s|\epsilon,b,n)ds = \frac{\epsilon^\alpha e^{-\epsilon s}(\epsilon s + b)^n s^{\alpha-1}}{\Gamma(\alpha+n)M(-n, 1-n-\alpha, b)} ds$$

6.2 Posterior for ϵ of the subsidiary measurement

The subsidiary measurement observes an integer number of events m, Poisson distributed as:

$$P(m|\epsilon) = \frac{e^{-\kappa\epsilon}(\kappa\epsilon)^m}{m!}$$

where κ is real number (connecting the subsidiary measurement to the main measurement) whose uncertainty is negligible, so κ can safely be treated as a fixed constant. κ might be thought of, for example, as based on a cross section that is exactly calculable by theory. There is negligible (i.e. zero) background in the subsidiary measurement.

The prior for ϵ is specified to be flat. The Bayesian posterior p.d.f. for ϵ is then

$$p(\epsilon|m) = \frac{\kappa(\kappa\epsilon)^m e^{-\kappa\epsilon}}{m!}$$

(or $\Gamma(m+1)$ instead of $m!$ in the denominator if you prefer). This is known as a gamma distribution.

The mean and rms of this posterior p.d.f. summarize the result of the subsidiary measurement as:

$$\epsilon = \frac{m+1}{\kappa} \pm \frac{\sqrt{m+1}}{\kappa} = \epsilon_0 \pm \sigma_\epsilon$$

Note that the observed data quantity in the subsidiary measurement is an integer m, while the quantity being measured by the subsidiary measurement is a positive real number ϵ.

6.3 Posterior for s with gamma prior for ϵ (b fixed)

Next we compute the joint posterior $p(s,\epsilon|b,n)dsd\epsilon$ using the $s^{\alpha-1}$ prior for s and our gamma distribution prior (i.e. the posterior derived from the subsidiary measurement) for $\epsilon \geq 0$

$$\text{prior for } \epsilon\text{:} \quad \pi(\epsilon)d\epsilon = \frac{(\kappa\epsilon)^\mu e^{-\kappa\epsilon}}{\Gamma(\mu)}\frac{d\epsilon}{\epsilon} \qquad \mu = m+1 = (\epsilon_0/\sigma_\epsilon)^2 \qquad \kappa = \epsilon_0/{\sigma_\epsilon}^2$$

where it is convenient to write μ for $m+1$. We have for the joint posterior p.d.f.

$$p(s,\epsilon|b,n)dsd\epsilon = \frac{1}{\mathcal{N}_{s,\epsilon}}\pi(\epsilon)e^{-\epsilon s}(\epsilon s+b)^n s^{\alpha-1}dsd\epsilon$$

where

$$\mathcal{N}_{s,\epsilon} = \int_0^\infty\int_0^\infty \pi(\epsilon)e^{-\epsilon s}(\epsilon s+b)^n s^{\alpha-1}dsd\epsilon = \int_0^\infty \pi(\epsilon)\mathcal{N}_s d\epsilon$$

We calculated $\mathcal{N}_s$ above, so we have

$$\mathcal{N}_{s,\epsilon} = \Gamma(\alpha+n)M(-n,1-n-\alpha,b)\int_0^\infty \epsilon^{-\alpha}\pi(\epsilon)d\epsilon$$

$$\mathcal{N}_{s,\epsilon} = \kappa^\alpha\Gamma(\mu-\alpha)\Gamma(\alpha+n)M(-n,1-n-\alpha,b)/\Gamma(\mu)$$

$$p(s,\epsilon|b,n)dsd\epsilon = \frac{\kappa^{\mu-\alpha}\epsilon^{\mu-1}s^{\alpha-1}(\epsilon s+b)^n e^{-(s+\kappa)\epsilon}}{\Gamma(\mu-\alpha)\Gamma(\alpha+n)M(-n,1-n-\alpha,b)}dsd\epsilon$$

The marginalized posterior for s can then be expressed as

$$p(s|b,n)ds = \left[\int_0^\infty p(s,\epsilon|b,n)d\epsilon\right]ds = \frac{s^{\alpha-1}\kappa^{\mu-\alpha}\mathcal{I}_\epsilon}{\Gamma(\mu-\alpha)\Gamma(\alpha+n)M(-n,1-n-\alpha,b)}ds$$

where the integral $\mathcal{I}_\epsilon$ is given by

$$\mathcal{I}_\epsilon = \int_0^\infty \epsilon^{\mu-1} e^{-(s+\kappa)\epsilon} (\epsilon s + b)^n d\epsilon$$

The same procedure that was used for the normalization integral can be applied here, producing

$$\mathcal{I}_\epsilon = \frac{b^{\mu+n}}{s^\mu} \int_0^\infty u^{\mu-1} e^{-b(1+\kappa/s)u} (1+u)^n du$$

$$\mathcal{I}_\epsilon = \frac{s^n n!}{(s+\kappa)^{\mu+n}} \sum_{k=0}^{n} \frac{\Gamma(\mu+n-k)}{(n-k)!k!} \left[\frac{b(s+\kappa)}{s}\right]^k$$

$$\mathcal{I}_\epsilon = \frac{s^n}{(s+\kappa)^{\mu+n}} \Gamma(\mu+n) M(-n, 1-n-\mu, b(s+\kappa)/s)$$

$$p(s|b,n)ds = \frac{\Gamma(\mu+n)}{\Gamma(\mu-\alpha)\Gamma(\alpha+n)} \frac{s^{\alpha+n-1}\kappa^{\mu-\alpha}}{(s+\kappa)^{\mu+n}} \frac{M(-n, 1-n-\mu, b(s+\kappa)/s)}{M(-n, 1-n-\alpha, b)} ds$$

which has a particularly simple form when the background term is zero:

$$p(s|b=0,n)ds = \frac{\Gamma(\mu+n)}{\Gamma(\mu-\alpha)\Gamma(\alpha+n)} \frac{s^{\alpha+n-1}\kappa^{\mu-\alpha}}{(s+\kappa)^{\mu+n}} ds$$

a Beta distribution of the 2nd kind. Note that we must require $\mu > \alpha > 0$ to obtain a normalizable posterior.

Our posterior p.d.f. for s with ϵ (and b) fixed is recovered exactly by taking the limit of $p(s|b,n)$ as $\sigma_\epsilon \to 0$. This means that the limit of s_u as $\sigma_\epsilon \to 0$ is identical to the value of s_u when ϵ is known exactly. This property may seem obvious, but it is violated by some frequentist methods of setting limits, so it is worth mentioning.

6.4 Calculating the limit

We need to integrate $p(s|b,n)$ up to some limit s_u, which can be done analytically as follows.

$$\int_0^{s_\mathrm{u}} p(s|b,n)ds = \frac{\Gamma(\mu+n)}{\Gamma(\mu-\alpha)\Gamma(\alpha+n)} \int_0^{\frac{s_\mathrm{u}}{s_\mathrm{u}+\kappa}} t^{\alpha+n-1}(1-t)^{\mu-\alpha-1} \frac{M(-n, 1-n-\mu, b/t)}{M(-n, 1-n-\alpha, b)} dt$$

where the substitution $t = \frac{s}{s+\kappa}$ has been performed. Re-expanding the polynomial M and integrating term by term yields

$$\int_0^{s_\mathrm{u}} p(s|b,n)ds = \sum_{k=0}^{n} \frac{I_x(\alpha+n-k, \mu-\alpha) n^{\underline{k}}}{(\alpha+n-1)^{\underline{k}}} \frac{b^k}{k!} \Bigg/ \sum_{k=0}^{n} \frac{n^{\underline{k}}}{(\alpha+n-1)^{\underline{k}}} \frac{b^k}{k!} \qquad \left(x = \frac{s_\mathrm{u}}{s_\mathrm{u}+\kappa}\right)$$

where I_x is the standard notation for the Incomplete Beta Function

$$\text{incomplete beta function:} \qquad I_x(q,r) \equiv \frac{\Gamma(q+r)}{\Gamma(q)\Gamma(r)} \int_0^x t^{q-1}(1-t)^{r-1} dt$$

which also satisfies the following recursion:

$$I_x(q,r) = \frac{\Gamma(q+r)}{\Gamma(q+1)\Gamma(r)} x^q (1-x)^r + I_x(q+1, r)$$

6.5 Integer moments of the marginalized posterior

Using the same technique as above, we can calculate the jth moment of the posterior p.d.f. as

$$\langle s^j \rangle = \int_0^\infty s^j p(s|b,n)ds = \frac{(\alpha+n)^{\overline{j}}\kappa^j}{(\mu-\alpha-1)^{\underline{j}}} \frac{M(-n,1-n-\alpha-j,b)}{M(-n,1-n-\alpha,b)}$$

where we utilize the notation[4]

$$\text{rising factorial:} \qquad z^{\overline{k}} \equiv \frac{\Gamma(z+k)}{\Gamma(z)} = z(z+1)(z+2)\cdots(z+k-1)$$

The expression for the mean of the posterior when $\alpha = 1$ can be simplified using the identity

$$M(-n,-n-1,b) = \left(1 - \frac{b}{n+1}\right) M(-n,-n,b) + \frac{b^{n+1}}{(n+1)!}$$

obtaining

$$\text{mean}(\alpha = 1) = \langle s \rangle|_{\alpha=1} = \frac{\kappa(n+1-b)}{\mu-2} + \frac{\kappa b^{n+1}}{(\mu-2)n!M(-n,-n,b)}$$

Note that the 2nd term is very small when $n \gg b$.

The recurrence relation[5]

$$r(r-1)M(q,r-1,z) + r(1-r-z)M(q,r,z) + z(r-q)M(q,r+1,z) = 0$$

leads to a recurrence relation between moments

$$\langle s^j \rangle = \frac{\kappa(\alpha+n+j-1-b)}{\mu-\alpha-j}\langle s^{j-1}\rangle + \frac{\kappa^2 b(\alpha+j-2)}{(\mu-\alpha-j+1)(\mu-\alpha-j)}\langle s^{j-2}\rangle$$

The special case $\alpha = 1$ then yields

$$\langle s^2 \rangle|_{\alpha=1} = \frac{\kappa^2}{(\mu-2)(\mu-3)}\left[(2+n-b)(1+n-b) + b + \frac{(2+n-b)b^{n+1}}{n!M(-n,-n,b)}\right]$$

which leads to this approximation for the variance of the posterior

$$\text{variance}(\alpha = 1) \simeq \frac{\kappa^2(1+n)}{(\mu-2)(\mu-3)} + \frac{\kappa^2(1+n-b)^2}{(\mu-2)^2(\mu-3)} \qquad (n \gg b)$$

6.6 Posterior for s with gamma priors for ϵ and b

Here we very briefly consider the case where the background parameter b also acquires an uncertainty. This case is more general than the fixed b case that is the main subject of this note: The fixed b case will be the subject of additional studies employing various popular frequentist techniques, with the goal of comparing their performance.

We judge the more general case considered in this subsection to be more complicated than necessary for the purpose of comparing the various methods, but it is instructive to document the fact that the Bayesian method can easily handle the more general case.

We assume a 2nd subsidiary measurement observing r events (Poisson, as was the case for ϵ), which, when combined with a flat prior for b, results in a gamma posterior for b of the form

$$p(b|r)db = \frac{\omega(\omega b)^r e^{-\omega b}}{r!} db$$

where ω is a calibration constant (analogous to κ in the subsidiary measurement for ϵ).

The posterior for b becomes the prior for b in the measurement of s. After determining the joint posterior $p(s, \epsilon, b|n)$ by using our priors for s, ϵ and b, we marginalize with respect to ϵ and b, resulting in

$$p(s|n)ds = \frac{\Gamma(\mu+n)}{\Gamma(\mu-\alpha)\Gamma(\alpha+n)} \frac{s^{\alpha+n-1}\kappa^{\mu-\alpha}}{(s+\kappa)^{\mu+n}} \frac{F(-n,\rho;1-n-\mu;(s+\kappa)/(s\omega))}{F(-n,\rho;1-n-\alpha;1/\omega)} ds$$

where we write $\rho = r+1$ for convenience, and F is the hypergeometric function[6]. As long as n is a non-negative integer and $\alpha > 0$, $F(-n, \rho; 1-n-\alpha; x)$ is a polynomial of order n in x (closely related to Jacobi polynomials).

This marginalized posterior for s can be then be integrated, with the result

$$\int_0^{s_\mathrm{u}} p(s|n)ds = \sum_{k=0}^{n} \frac{I_x(\alpha+n-k,\mu-\alpha)n^{\underline{k}}\rho^{\overline{k}}}{(\alpha+n-1)^{\underline{k}}} \frac{\omega^{-k}}{k!} \Bigg/ \sum_{k=0}^{n} \frac{n^{\underline{k}}\rho^{\overline{k}}}{(\alpha+n-1)^{\underline{k}}} \frac{\omega^{-k}}{k!} \qquad \left(x = \frac{s_\mathrm{u}}{s_\mathrm{u}+\kappa}\right)$$

These two equations closely resemble the main results of sections 6.3 and 6.4: to recover the fixed b results, simply substitute $b\omega$ for ρ above, and take the limit $\omega \to \infty$.

7 Appendix B—Average Coverage Theorem

In this appendix we prove that Bayesian credible intervals have average frequentist coverage, where the average is calculated with respect to the prior density. We start from the Bayesian posterior density:

$$p(s\,|\,n) \;=\; \frac{P(n\,|\,s)\,\pi(s)}{\int_0^\infty P(n\,|\,s)\,\pi(s)\,ds}. \tag{13}$$

For a given observed value of n, a credibility-β Bayesian interval for s is any interval $[s_\mathrm{L}(n), s_\mathrm{U}(n)]$ that encloses a fraction β of the total area under the posterior density. Such an interval must therefore satisfy:

$$\beta \;=\; \int_{s_\mathrm{L}(n)}^{s_\mathrm{U}(n)} p(s\,|\,n)\,ds, \tag{14}$$

or, using the definition of the posterior density:

$$\int_{s_\mathrm{L}(n)}^{s_\mathrm{U}(n)} P(n\,|\,s)\,\pi(s)\,ds \;=\; \beta \int_0^\infty P(n\,|\,s)\,\pi(s)\,ds. \tag{15}$$

Now for coverage. Given a true value s_t of s, the coverage $C(s_\mathrm{t})$ of $[s_\mathrm{L}(n), s_\mathrm{U}(n)]$ is the frequentist probability that s_t is included in that interval. We can write this as:

$$C(s_\mathrm{t}) \;=\; \sum_{\substack{n \text{ such that:}\\ s_\mathrm{L}(n)\le s_\mathrm{t}\le s_\mathrm{U}(n)}} P(n \,|\, s_\mathrm{t}). \tag{16}$$

Next we calculate the average coverage $\overline{C}$, weighted by the prior $\pi(s)$:

$$\begin{aligned}
\overline{C} &= \int_0^\infty C(s)\,\pi(s)\,ds, \\
&= \int_0^\infty \sum_{\substack{n \text{ such that:}\\ s_\mathrm{L}(n)\le s\le s_\mathrm{U}(n)}} P(n \,|\, s)\,\pi(s)\,ds\;, && \text{using equation (16),} \\
&= \sum_{n=0}^\infty \int_{s_\mathrm{L}(n)}^{s_\mathrm{U}(n)} P(n \,|\, s)\,\pi(s)\,ds, && \text{interchanging}^{12} \text{ integral and sum,} \\
&= \beta \sum_{n=0}^\infty \int_0^\infty P(n \,|\, s)\,\pi(s)\,ds, && \text{using equation (15),} \\
&= \beta \int_0^\infty \sum_{n=0}^\infty P(n \,|\, s)\,\pi(s)\,ds, && \text{interchanging sum and integral,} \\
&= \beta \int_0^\infty \pi(s)\,ds, && \text{by the normalization of } P(n \,|\, s), \\
&= \beta, && \text{by the normalization of } \pi(s).
\end{aligned}$$

This completes the proof.

[12]This step is not trivial; some care is required to perform this correctly.

References

[1] M. J. Bayarri and J. O. Berger, "The Interplay of Bayesian and Frequentist Analysis", Statistical Science 19, 58 (2004), projecteuclid.org/Dienst/UI/1.0/Summarize/euclid.ss/1089808273 www.isds.duke.edu/~berger/papers/interplay.html

[2] J. M. Bernardo and A. F. M. Smith, "Bayesian Theory", (John Wiley and Sons, Chichester, UK, 1993), §5.4 and §A.2.

[3] Harold Jeffreys, "Theory of Probability", 3rd ed., (Oxford University Press, Oxford, 1961), §3.1.

[4] R. L. Graham, D. E. Knuth, and O. Patashnik, "Concrete Mathematics: A Foundation for Computer Science", 2nd ed., (Addison-Wesley, Reading, MA, 1994); planetmath.org/encyclopedia/FallingFactorial.html.

[5] M. Abramowitz and I.A. Stegun, editors, "Handbook of Mathematical Functions", (United States Department of Commerce, National Bureau of Standards, Washington, D.C. 1964; and Dover Publications, New York, 1968), chapter 13.

[6] M. Abramowitz and I.A. Stegun, ibid., chapter 15; William H. Press, et al., "Numerical Recipes", 2nd edition, (Cambridge University Press, Cambridge, 1992), §5.14 and §6.12 lib-www.lanl.gov/numerical/bookcpdf/c5-14.pdf lib-www.lanl.gov/numerical/bookcpdf/c6-12.pdf

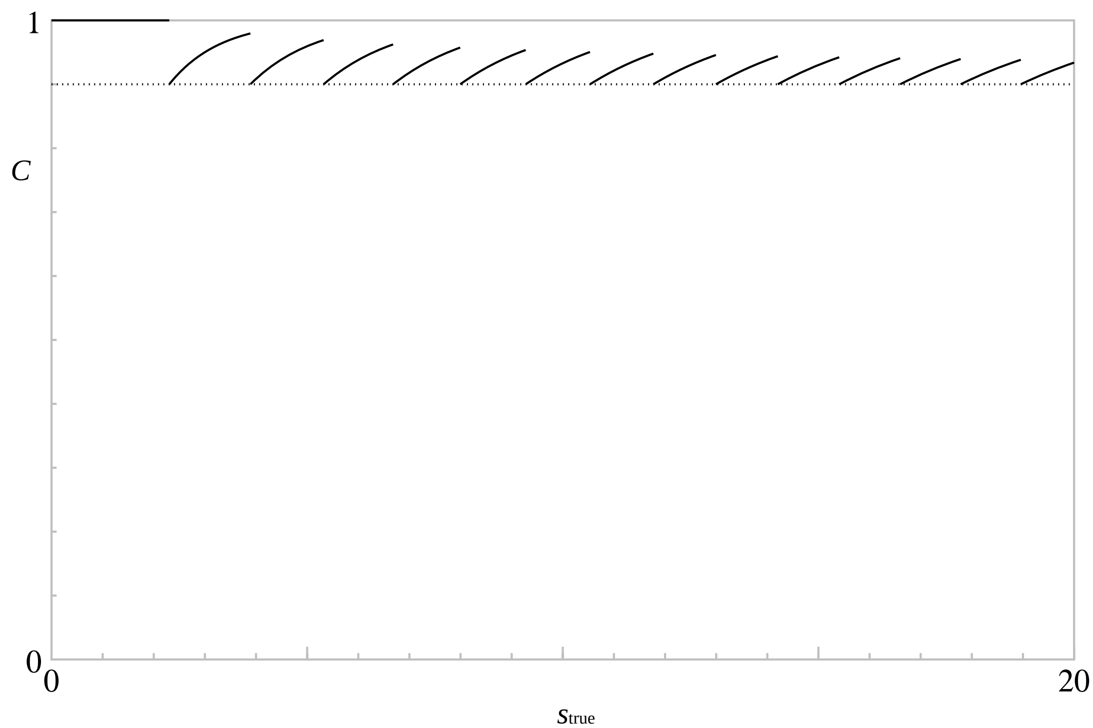

Figure 1: Coverage as a function of the true signal rate s for Bayes 90% limits, for the simple case of no background and no uncertainty on $\epsilon = 1$.

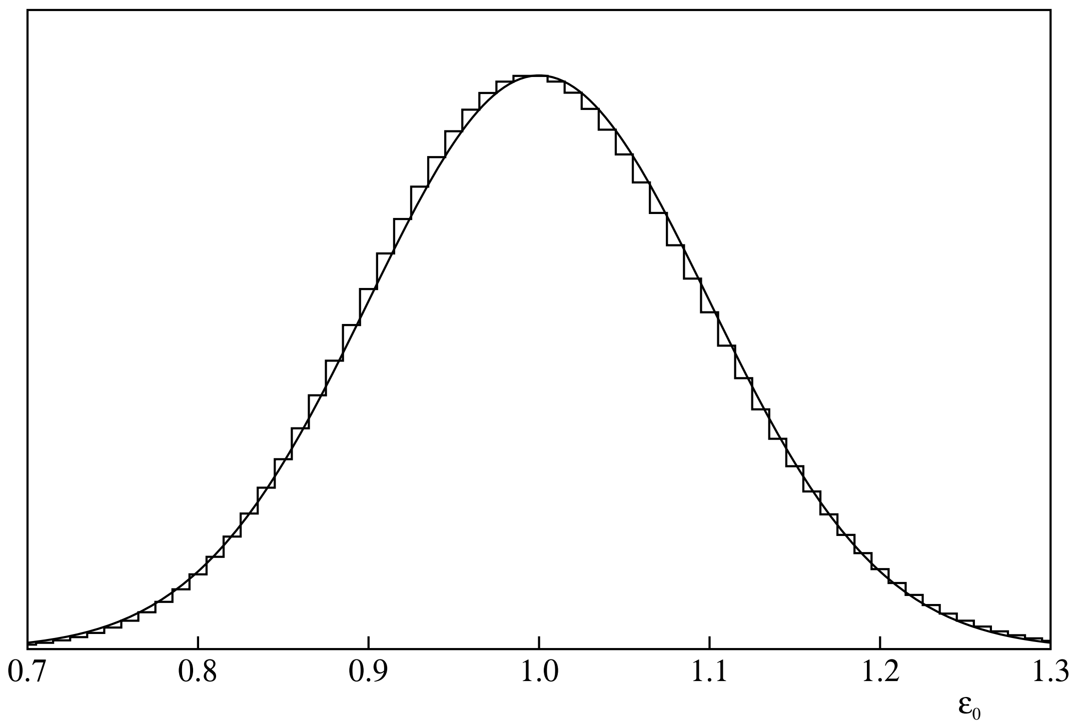

Figure 2: Comparison of our discrete probability for ϵ_0 (shown as a histogram, see eqn (10)) and Gaussian (continuous curve) for the case $\epsilon = 1 \pm 0.1$.

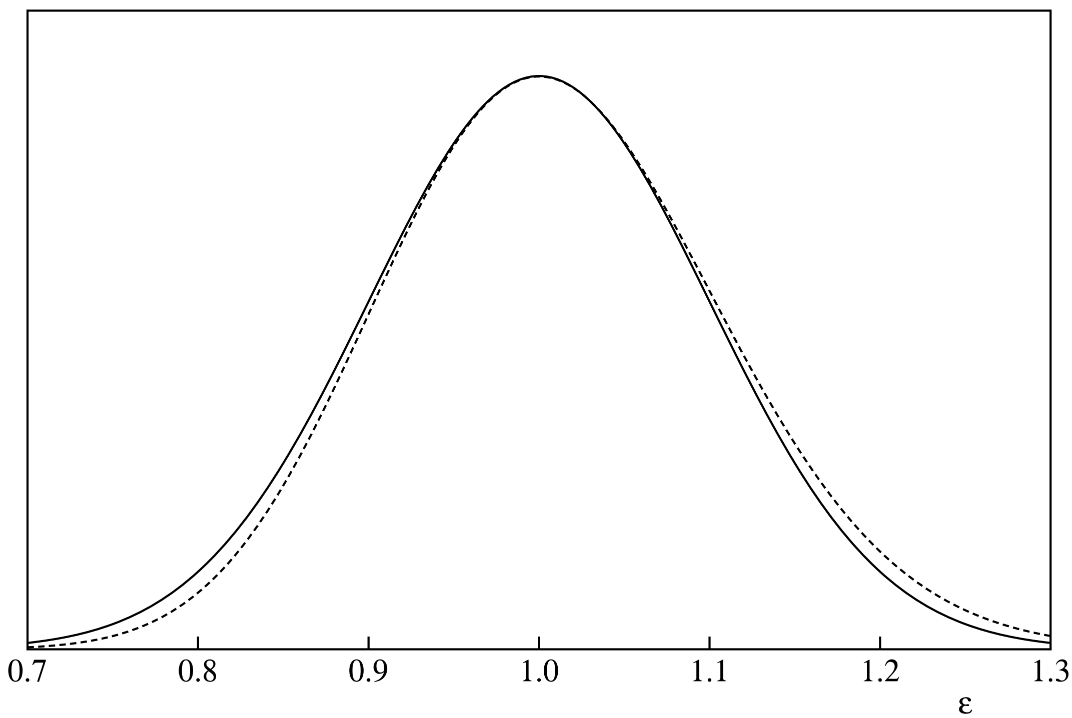

Figure 3: Comparison of our likelihood (dashed, see eqn (11)) and Gaussian (solid) for the case $\epsilon = 1 \pm 0.1$.

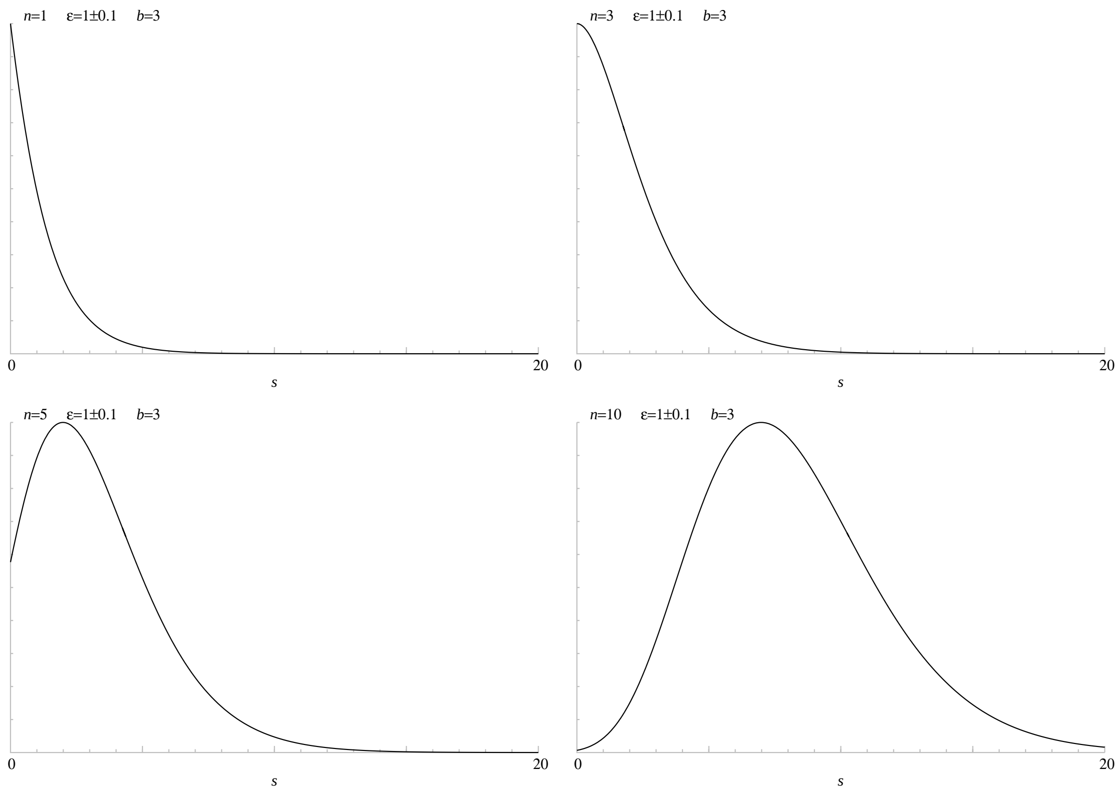

Figure 4: Posterior densities $p(s|b,n)$ vs s for $n = 1$, 3, 5, 10. In each case, $b = 3$ and $\epsilon = 1 \pm 0.1$ (i.e. $\kappa = 100$ and m=99).

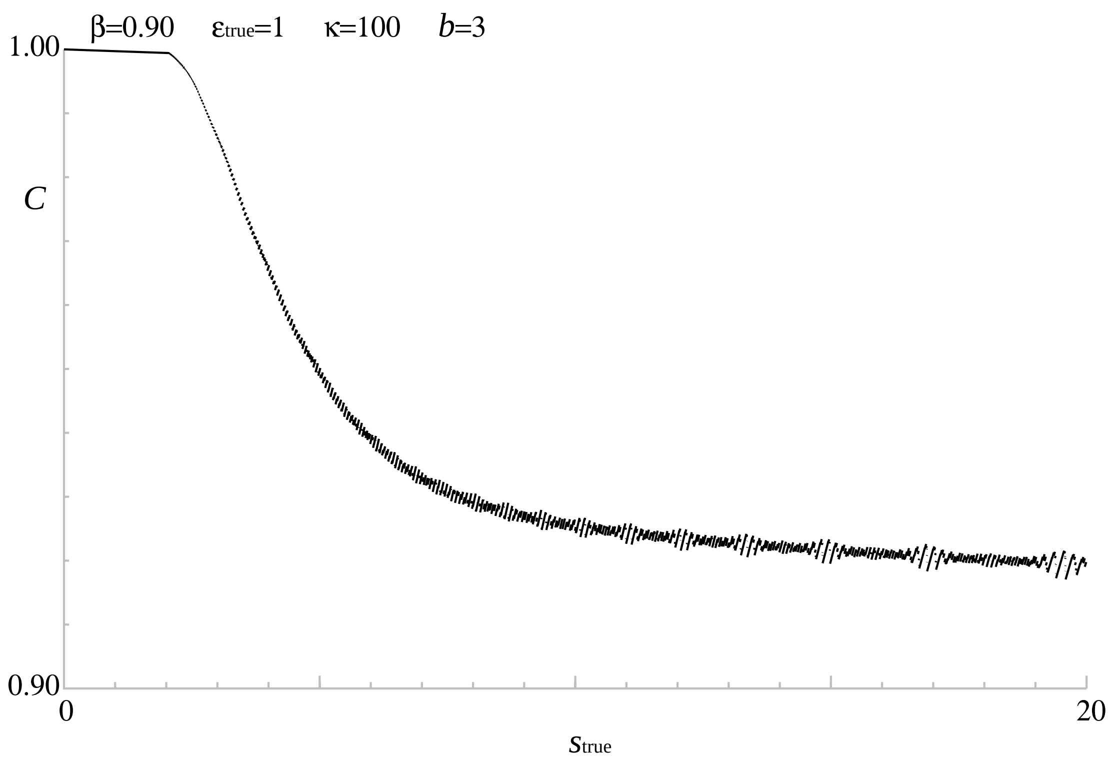

Figure 5: Coverage of 90% upper limits as a function of s_{true} for $\epsilon_{\mathrm{true}} = 1$, nominal 10% uncertainty of the subsidiary measurement of ϵ, and $b = 3$ background expected.

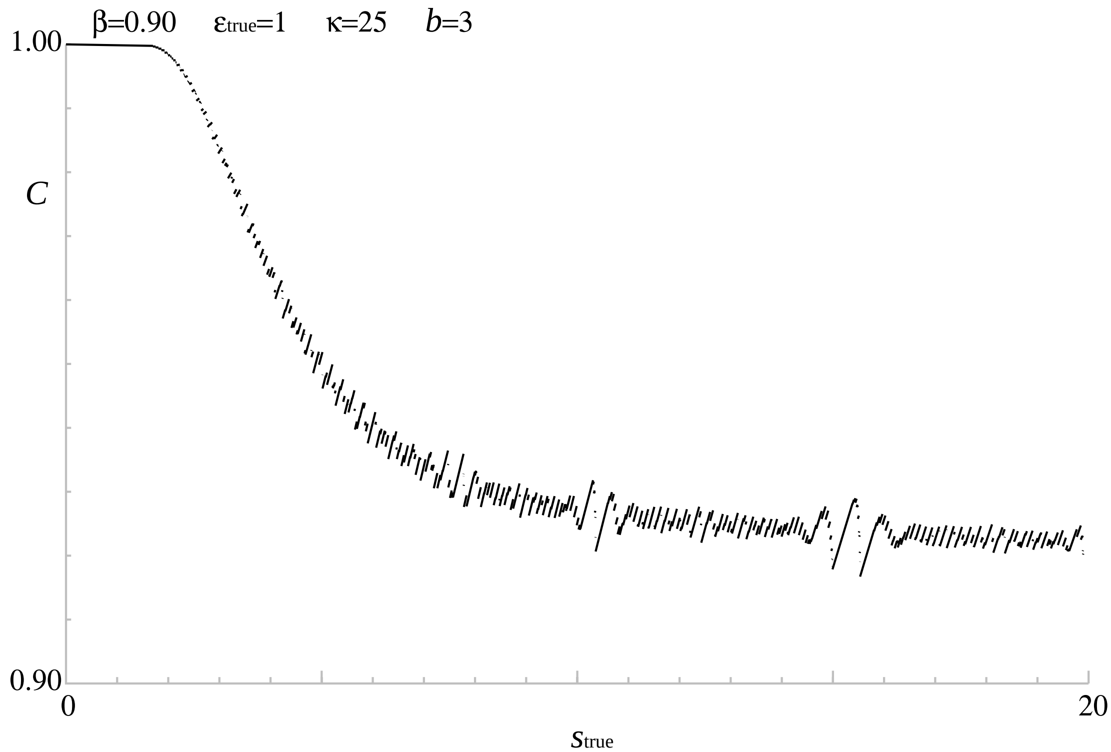

Figure 6: Coverage of 90% upper limits as a function of s_{true} for $\epsilon_{\text{true}} = 1$, nominal 20% uncertainty of the subsidiary measurement of ϵ, and $b = 3$ background expected.

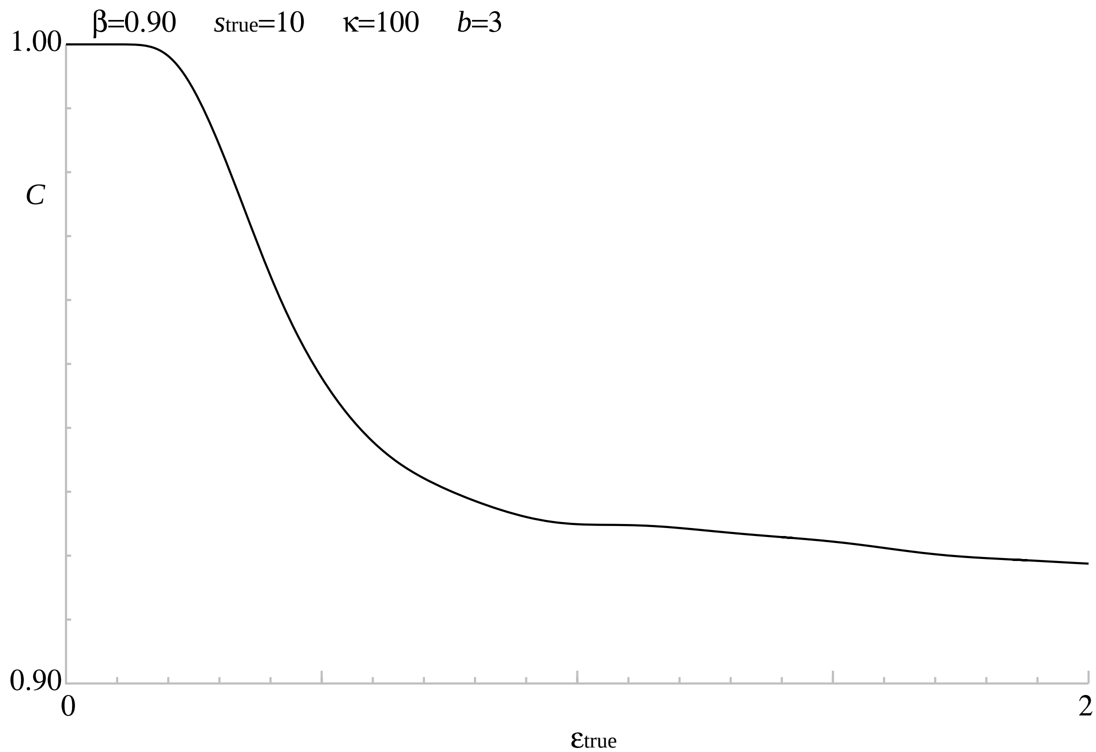

Figure 7: Coverage of 90% upper limits as a function of $\epsilon_{\rm true}$ for $s_{\rm true} = 10$, nominal 10% uncertainty of the subsidiary measurement of ϵ, and $b = 3$ background expected.

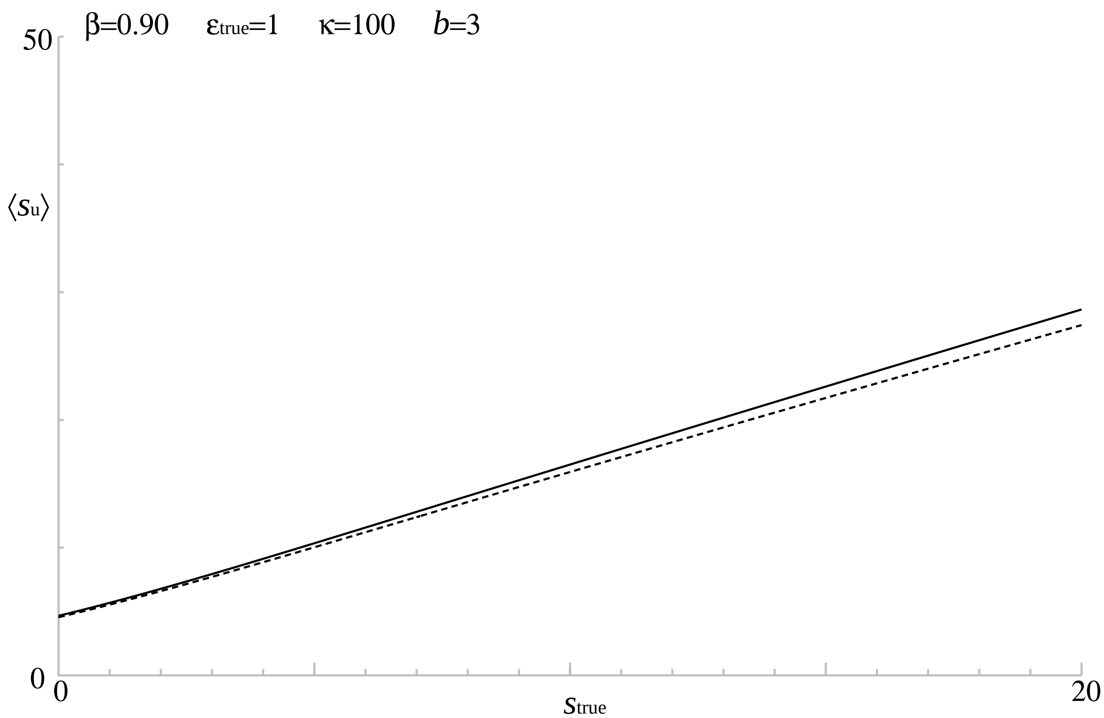

Figure 8: Sensitivity of 90% upper limits as a function of $s_{\rm true}$ for $\epsilon_{\rm true} = 1$, nominal 10% uncertainty of the subsidiary measurement of ϵ, and $b = 3$ background expected. For reference, the sensitivity for $\sigma_\epsilon = 0$ is also given (dashed).